Appraisal of candidate instruments for assessment of the physical function domain in patients with psoriatic arthritis

Ying Ying Leung[1], Ana-Maria Orbai[2], Alexis Ogdie[3], Pil Hojgaard[4], Richard Holland[5], Niti Goel[6], Jeffrey Chau[7], Laura C Coates[8], Vibeke Strand[9], Dafna D Gladman[10], Philip Mease[11], Robin Christensen[4, 12], William Tillett[13]

1. Singapore General Hospital, Duke-NUS Medical School, Singapore
2. Director Psoriatic Arthritis Program, Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA
3. Medicine and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA
4. Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Denmark
5. Concord Repatriation General Hospital, Sydney, Australia
6. Patient Research Partner, Adjunct Assistant Professor, Duke University School of Medicine, Durham, North Carolina, USA
7. Patient Research Partner, Hong Kong Psoriatic Arthritis Association, Hong Kong, China
8. National Institute for Health Research Clinician Scientist, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom
9. Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA
10. Division of Rheumatology, University of Toronto, Senior Scientist, Krembil Research Institute, Director, Psoriatic Arthritis Program, University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada
11. Swedish Medical Center/Providence St Joseph Health and University of Washington School of Medicine, Seattle, Washington, USA
12. Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark.
13. Royal National Hospital for Rheumatic Diseases, University of Bath, Bath, United Kingdom

Correspondence to: Ying-Ying Leung, MD; Department of Rheumatology and Immunology, Singapore General Hospital, The Academia, level 4, 20 College Road, Singapore 169856,

Contact No.: +65 63265276, Fax no.: +65 62203321, E-mail: katyccc@hotmail.com

ABSTRACT

Objective. Numerous Patient-Reported Outcome Measures (PROMs) exist for the measurement of physical function for psoriatic arthritis (PsA), but only a few are validated. The objective of this project was to prioritize PROMs for measuring physical function for potential incorporation into a standardized Outcome Measurement Set for PsA

Methods. A working group of 13 members including two patient research partners was formed. We applied a template to assess and prioritize evaluation of PROMs for physical function, through their identification, discussions and Delphi exercises to achieve consensus.

Results. PROMs measuring physical function in PsA were identified through a systematic literature review and recommendations by the working group. The rationale for inclusion and exclusion from the original list of existing PROMs was thoroughly discussed and two rounds of Delphi exercises were conducted to achieve consensus. Six PROMs were prioritized: Health Assessment Questionnaire and four modifications (HAQ-Disability Index, HAQ-Spine, modified HAQ, Multidimensional HAQ), Medical Outcome Survey Short Form-36 (SF-36)-physical functioning domain and the PROMIS physical functioning module.

Conclusion. We prioritized six physical function PROMs for PsA. These six PROMs will undergo further appraisal using the Outcome Measures in Rheumatology (OMERACT) Filter 2.1.

Key Indexing Terms: Psoriatic Arthritis, Psoriasis, Outcome Measures, Physical Function

Source of Support: YYL is funded by the Clinician Scientist award of the National Medical Research Council, Singapore (NMRC/CSA-INV/0022/2017). The views expressed are those of the author(s) and not necessarily those of the NMRC. AMO is funded by the Jerome L. Greene Foundation Scholar Award, the Staurulakis Family Discovery Award, the Rheumatology Research Foundation, and the National Institutes of Health (NIH) through the Rheumatic Diseases Resource-based Core Center (P30-AR053503 Cores A and D, and P30-AR070254, Cores A and B). All statements in this report, including its findings and conclusions, are solely those of the authors and do not necessarily represent the views of the NIH, the National Institute of Arthritis Musculoskeletal and Skin Diseases (NIAMS), or the Rheumatology Research Foundation (RRF). PH and RC (The Parker Institute, Bispebjerg and Frederiksberg Hospital) is supported by a core grant from the Oak Foundation (OCAY-13-309). LCC is funded by a National Institute for Health Research Clinician Scientist award. The research was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). The views expressed are those of the author(s) and not necessarily those of the NHS, the NIHR, or the Department of Health. AO is funded by the Rheumatology Research Foundation and NIH/NIAMS R01 AR072363.

Author Information: YY Leung, MB ChB, MD, [ORCID: 0000-0001-8492-6342], Associate Professor, Duke-NUS Medical School, Singapore, Department of Rheumatology and Immunology, Singapore General Hospital, Singapore, katyccc@hotmail.com; A-M Orbai, MD, MHS, [ORCID: 0000-0001-8644-8567], Assistant Professor of Medicine, Director Psoriatic Arthritis Program, Division of Rheumatology, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA, aorbai1@jhmi.edu; A Ogdie, MD, MSCE, [ORCID: 0000-0002-4639-0775], Associate Professor of Medicine and Epidemiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA, alogdie@pennmedicine.upenn.edu; P Hojgaard, MD, PhD, [ORCID: 0000-0002-8046-263X],

Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Denmark, pil.hoejgaard.01@regionh.dk; R Holland, MBBS, [ORCID: 0000-0001-7227-8210], Concord Repatriation General Hospital, Sydney, Australia, drrichardholland@gmail.com; N Goel, MD, [ORCID: 0000-0001-5869-5157], Patient Research Partner, Adjunct Assistant Professor, Duke University School of Medicine, Durham, North Carolina, USA, niti.goel@caduceusbio.com; J Chau, BBA, [ORCID: 0000-0002-3484-7977], patient research partner, yf.jeffchau@gmail.com; LC Coates, MB ChB, PhD, [ORCID: 0000-0002-4756-663X], National Institute for Health Research Clinician Scientist, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford, United Kingdom, laura.coates@ndorms.ox.ac.uk; V Strand, MD, [ORCID: 0000-0003-4978-4072], Division of Immunology/Rheumatology, Stanford University School of Medicine, Palo Alto, California, USA, vstrand@stanford.edu; DD Gladman, MD, FRCPC, [ORCID: 0000-0002-9074-0592], Professor of Medicine, University of Toronto, Senior Scientist, Krembil Research Institute, Director, Psoriatic Arthritis Program, University Health Network, Toronto Western Hospital, Toronto, Ontario, Canada, dafna.gladman@utoronto.ca; PJ Mease, MD, [ORCID: 0000-0002-6620-0457], Rheumatology Research, Swedish Medical Center and University of Washington School of Medicine, Seattle, Washington, USA, pmease@philipmease.com; R Christensen, PhD, [ORCID: 0000-0002-6600-0631], Musculoskeletal Statistics Unit, The Parker Institute, Bispebjerg and Frederiksberg Hospital, Denmark, Research Unit of Rheumatology, Department of Clinical Research, University of Southern Denmark, Odense University Hospital, Denmark, robin.christensen@regionh.dk; W Tillett, BSc, MB ChB, PhD, MRCP, [ORCID: 0000-0001-7531-4125], Consultant Rheumatologist, Senior Lecturer, Royal National Hospital for Rheumatic Diseases, Pharmacy and Pharmacology, University of Bath, Bath, United Kingdom, w.tillett@nhs.net.

Running Footline: Shortlisting PROMs for physical function in PsA

Acknowledgement:

We would like to thank Dorcus Beaton and Lara L. Maxwell for technical advice and review of this manuscript.

INTRODUCTION

Psoriatic arthritis (PsA) is a chronic inflammatory disease with manifestations including arthritis, enthesitis, dactylitis, spondylitis, skin and nail psoriasis (1, 2). PsA causes damage of articular joints and can profoundly impact physical function and health-related quality of life (HRQoL) in affected individuals. The Group for Research and Assessment of Psoriasis and Psoriatic Arthritis (GRAPPA) and Outcome Measures in Rheumatology (OMERACT) are working to combine perspectives of care providers, researchers and patient research partners (PRPs) to update the PsA Core Outcome Set which identifies the key outcomes to be measured in randomized controlled trials (RCTs) and longitudinal observational studies (LOS) (3). Core Outcome Sets represent the minimum domains that should be measured and reported in all RCTs and LOS of a specific condition (4). Use of Core Outcome Sets does not imply that outcomes in a particular RCT should be restricted to those endpoints. OMERACT advocates that each trial should measure the Core Outcome Set which is based on both a Core Domain Set (the *What* to measure) and a Core Outcome Measurement Set (the *How* to measure) (5). A Core Domain Set for PsA was updated and endorsed in 2016 (3).

The lack of standardization of outcome measurement instruments in PsA RCTs and LOS has been highlighted, resulting in inconsistency of data reporting and heterogeneity in results (6). After finalizing the Core Domain Set, the GRAPPA-OMERACT PsA Core Outcome Set working group is currently leading the effort to develop and ratify a standardized Core Outcome Measurement Set (7). The process to do so follows recommendations outlined in the OMERACT (Outcome Measurement in Rheumatology) Filter 2.1 (5, 8). The OMERACT Filter 2.1 is a set of standards for evidence-based decision making which addresses Core Outcome Set development. Endorsing a measurement

instrument to assess a certain domain using the OMERACT Filter 2.1 involves multiple work streams including systematic literature reviews (SLR), with appraisal and synthesis of the evidence on instrument properties; discussions among stakeholders, and Delphi consensus exercises. The synthesis of evidence follows the pillars of OMERACT Filter 2.1: Domain match (instrument measuring what it is supposed to measure), Feasibility (instrument is practical to use), Truth (degree to which the instrument's score make numerical sense) and Discrimination (instrument can distinguish situation of no change versus change, is sensitive to change in RCTs, and has a threshold of meaning for interpretation) (5).

Physical function is included in the PsA Core Domain Set as it has been identified as one of the core domains reflecting disease impact in PsA patients (9-11). Several instruments are available to measure physical function in PsA, including those originally developed for use in other conditions, such as rheumatoid arthritis (RA), as well as newer instruments developed specifically for PsA (12). The process to prioritize instruments prior to appraisal using the OMERACT Filter 2.1 is conducted by individual working groups. The PsA Core Outcome Set working group steering committee developed a template to facilitate this process, and this template has been described elsewhere (13). It includes the formation of a working group, identification of instruments and preliminary appraisal of existing evidence, and discussions and Delphi exercises to prioritize instruments that have the highest potential to fulfill OMERACT Filter 2.1. This report details the steps taken by the physical function working group to prioritize patient-reported outcome measures (PROMs) for the assessment of the physical function domain in PsA that will be candidates for further consideration.

METHODS

This report describes application of a template to the physical function domain for PsA to prioritize instruments to undergo the OMERACT Filter 2.1.

1. Formation of a working group for the outcome domain.

The working group members were identified through GRAPPA, and included personnel with expertise in the physical function domain in PsA. Candidates were invited from within the steering committee and recommendation from working group members. The working group involved at least 2 Patient Research Partners (PRPs) who were invited to participate by the GRAPPA PRP Chair.

2. Identification and preliminary appraisal of measurement instruments for the domain.

We identified instruments for measuring physical function based on results from a recent systematic review of measurement properties of PROMs in PsA that involved both health professionals and PRPs (14). In the previous work, published articles with data regarding development or assessment of the measurement properties of PROMs were identified (14); these measurement properties were evaluated using the approach described by Prinsen et al (15) and the COnsensus-based Standards for the selection of health Measurement Instruments (COSMIN) checklist (16). The full process and results are described elsewhere (14). Each PROM was appraised for three main and eight subcategories, namely reliability (subcategories: internal consistency, test-retest reliability, measurement error), content validity (subcategories: content validity, structural validity, hypothesis testing, cross cultural validity, criterion validity), and responsiveness (16).

In addition, new and potential instruments that measure physical function were suggested by working group members.

3. Discussion and Delphi exercise to achieve concensus regarding instrument prioritization

A comprehensive document on the physical function PROMs was developed and presented to working group members (Supplementary document). This included the background of the PROMs, format and scoring methods. Included in the document was a Summary of Measurement Properties Table that detailed the measurement properties of the PROMs appraised in the previous work (14). A teleconference was conducted among working group members to discuss the various PROMs and the Delphi format. The working group decided on having two rounds of Delphi exercises, with interim discussions via teleconference or email to facilitate achieving consensus on prioritizing physical function PROMs. All Delphi exercises were conducted anonymously on online portals.

In the first Delphi exercise, working group members were asked to vote based on their own understanding of the PROMs. Working group members were advised to focus primarily on whether the PROMs matched to the domain of physical function in PsA and on the feasibility of the PROMs. Working group members were also given the Summary of Measurement Properties Table derived from the previous published work (14). However, this information was considered secondary, as the full set of evidence required by OMERACT Filter 2.1 had not been developed. In particular, RCT evidence for discrimination was not included. A question for each PROM was asked, "Do you think this PROM should be taken forward for further evaluation?". A simple yes/no response for each PROM was requested, and additional comments were collected as free text.

The results of the voting of the first Delphi exercise were discussed. The working group then drafted the questions for a second Delphi exercise. In the second Delphi exercise, it was prespecified that PROMs receiving ≥70% endorsement would be included for further appraisal using OMERACT Filter 2.1.

RESULTS

1. Formation of the physical function working group

A physical function working group of 13 members was formed in June 2018. The final members of the working group consisted of experts (10 rheumatologists and one methodologist) with experience in physical function measurement in PsA, and two PRPs. Working group members had international representation, spanning across three continents (countries of origin: Australia, Canada, Denmark, Hong Kong SAR of China, Singapore, UK, and USA). Two teleconference sessions with additional PRPs were conducted to explain the purpose of study, work flow, instruments for consideration of assessment of physical function domain and the OMERACT Filter 2.1 methodology.

2. Identification of PROMs for physical function

Physical function in PsA was defined as *"Being able to perform physical activities from daily to recreational activities (includes upper/lower extremity functioning, balance)"* (17). Examples of the concept of physical function were taken from quotations from a GRAPPA international focus group study (9) and summarized in Table 1 of the Supplementary document. Based on this definition and the concept of physical function being the perception of physical capability, the working group therefore decided to focus on PROMs instead of performance-based assessments.

We identified relevant physical function PROMs from a recent comprehensive SLR on PROMs for various domains in PsA (14). The evidence derived from the SLR for physical function PROMs was extracted from the published article (14) and presented to working group members for review and discussion (Supplementary document). These PROMs were: Health Assessment Questionnaire (HAQ)-Disability Index (-DI) (18), HAQ-Spondyloarthritis

(HAQ-S) (19), modified HAQ (mHAQ) (20), Physical Functioning domain of the Medical Outcomes Study 36-item Short Form Survey (SF-36 PF10) (21), Physical Component Summary Score of the SF-36 (SF-36 PCS) (21), PCS of the SF-12 (SF-12 PCS) (22), Psoriatic Arthritis Impact of Disease (PsAID) functional capacity item (23), Arthritis Impact Measurement Scales (AIMS) (24), Bath Ankylosing Spondylitis Functional Index (BASFI) (25), and the American College of Rheumatic Diseases (ACR) functional class (26). Two additional PROMs were suggested by working group members: multidimensional HAQ (MDHAQ) (27) and the Patient-Reported Outcomes Measurement Information System (PROMIS)-Short Form Physical Function 10a (PROMIS-PF10a) (28). The MDHAQ has been incorporated in the Routine Assessment of Patient Index Data 3 (RAPID3) that was developed for use in clinical care in RA (29), and is being incorporated as a routine measurement in clinical care for PsA in some countries. The PROMIS-PF10a was developed based on item banks for physical function.

Relevant information for these 12 physical function PROMs was summarized in a comprehensive document and circulated to all working group members (Supplementary document). The document listed all the 12 PROMs with a short introduction and scoring methods. A Summary of Measurement Properties table taken from the previous SLR of PROMs was enclosed. Working group members were asked to review the information and prepare for the Delphi exercise.

3. Working group discussions and Delphi exercises

The first Delphi exercise was conducted in June 2018 and finalized on 12 July 2018 via an anonymized online voting portal. All 13 working group members participated (response rate 100%). The voting results of the first Delphi exercise and comments made to various PROMs are summarized in Table 1.

The results of the first Delphi exercise were then presented to the working group members, followed by open discussion via email from 12 - 27 July 2018. A one-hour web-based discussion was conducted on 23 August 2018, followed by further open discussion via email from 23 August to 19 September 2018. During the teleconferences and subsequent e-mail communications, members of the working group were allowed to speak freely on their views of the PROMs. Based on the discussion points, a script for a second Delphi exercise was drafted and reviewed by all working group members. Several revisions of the phrasing and wording were done and finally agreed upon by all members of the working group (Table 2).

In the second round of the Delphi exercise, results of the overall voting of the working group in the first round of Delphi exercise and discussion points were made available. Again, working group members were asked whether or not to take the individual PROM to appraisal via OMERACT Filter 2.1, based on their understanding of domain match, feasibility and measurement properties. It was prespecified that instruments receiving <70% endorsement in the second Delphi exercise would be excluded from further formal appraisal. All 13 working group members participated in the second Delphi exercise and results with reasons for the inclusion or exclusion of all PROMs are summarized in Table 2.

The HAQ and modifications. The HAQ-DI was originally developed for RA and adapted for arthritis in general (18). It includes 20 items assessing eight aspects of physical function: dressing and grooming, arising, eating, walking, hygiene, reach, grip, and activities. As the most commonly used instrument to assess physical function in PsA RCTs (12), it received unanimous endorsement in both Delphi exercises.

The HAQ-S, a modification of the original HAQ-DI with 5 additional items assessing function of the axial spine, received only a 69% vote in the first Delphi. While analyses of

data have demonstrated that the HAQ-S does not capture additional information compared with HAQ-DI (30), some members thought that this result may have been related to the original PsA cohort in which the HAQ-S was tested and needed further testing in populations enriched for the presence of axial PsA. Both HAQ-DI and HAQ-S have been collected in the large Corrona registry in the United States, thus there is a potential that comparative data about performance of the two instruments in patients whose PsA includes axial involvement will become available from the registry. In the second Delphi exercise, use of the HAQ-S was addressed with two questions: the first was whether or not to include, and the second was to allow use of either the HAQ-DI or HAQ-S dependent on the clinical setting. With these considerations, the HAQ-S received 79% and 84% of the votes in favor of inclusion.

The mHAQ is a shortened version of HAQ-DI with only 8 items, one from each subdomain of the HAQ-DI (20); it received >70% of the votes for inclusion in both Delphi exercises. The MDHAQ, which includes the 8 items of mHAQ with 2 additional items (patient global assessment of disease activity and pain) (27), was presented as part of the RAPID3 in the first Delphi when it received only 69% of the votes. During the teleconference discussion, the 10-item MDHAQ was clarified as an instrument purely to assess physical function. Consensus was achieved to retain the MDHAQ in the second Delphi exercise, with a vote of 76% to be included.

The Medical Outcomes Study Surveys. The SF-36 PCS received 61.5% of the vote in the first Delphi. Although the results of SF-36 PCS scores have been reported in many RCTs, there were concerns expressed by the working group regarding the concept represented by the summary scores of the SF-36, as they are calculated based on positive and negative weighting of all 8 domains with a population norm of a mean (standard deviation) of 50 (10). The key utility of this norm-based scoring is for easy comparison of the summary scores at a group level with the normal population average scores in epidemiologic studies (31). However, the

SF-36 PCS represents a broader concept than physical function alone (21, 31), and therefore not having the domain match. The SF-36 PCS was excluded following the second Delphi exercise. In contrast, the PF domain of the SF-36 (SF-36 PF10) that includes 10 items measuring physical function matched with the domain intended to measure. The SF-36 PF10 received unanimous endorsement for inclusion from both Delphi exercises. It has been noted, however, that to use the SF-36 PF10, the entire SF-36 HRQoL questionnaire must be scored (21, 31).

Based on the same reasoning by which the SF-36 PCS was excluded, the working group felt the SF-12 PCS did not represent the physical function domain (lack of domain match), and there is no existing study that has evaluated its exclusive use in PsA. The SF-12 PCS was excluded from the second round of the Delphi exercise and further consideration.

The PsAID functional capacity item. PsAID is a PsA-specific derived multidimensional instrument that measures the life impact of PsA. It is often considered a HRQoL measure (23). Physical function is represented by a single item with an 11-point numeric rating scale (0-10) as functional capacity impact attributed to PsA. It received 84.6% of the votes in the first Delphi. Concerns were raised regarding the validity of utilizing a single item from a composite measure of HRQoL, and the domain match of the item itself. Consensus excluded the PsAID functional capacity item from further evaluation in the second Delphi exercise.

PROMIS-PF10a. Despite the lack of validation data, the working group thought that the PROMIS-PF10a could be a promising instrument. The PROMIS-PF10a was developed from a 1,728-item bank taken from 165 instruments assessing physical function. It received 92.3% and 100% of the votes for inclusion in the first and second Delphi exercises, respectively.

Other PROMs. The AIMS, BASFI and ACR functional class received 30.8%, 61.5% and 30.8% votes in the first Delphi exercise. Shortcomings for the AIMS include that it is too

long, thereby lacking feasibility, and it has not been used in the last decade. The BASFI was considered not to have adequate domain match as well as not providing additional information beyond the HAQ-DI. The ACR functional class was considered to be lacking domain match as it is too crude an instrument for measuring physical function in PsA patients who currently are less physically impaired or disabled following the new treatment strategies (32). These three instruments were considered as a single question in the second Delphi exercise and were excluded from further appraisal using the OMERACT Filter 2.1.

DISCUSSION

In this report we summarize the process leading to a preliminary prioritization of PROMs for assessment of physical function in PsA RCTs and LOS. Six PROMs for assessment of the physical function domain in PsA were successfully prioritized for further appraisal: HAQ-DI, HAQ-S, mHAQ, MDHAQ, SF-36 PF10, and PROMIS-PF10a. These prioritized PROMs will undergo formal appraisal of specific measurement properties using the OMERACT Filter 2.1 individually.

Members of GRAPPA are committed to standardizing the core outcome measurement set in PsA RCTs and LOS which is essential to minimize heterogeneity and facilitate interpretation of the studies (7). With updating of the PsA Core Outcome Set, research processes have been underway to evaluate instruments for each of the specified domains. We tested a consensus-based process for candidate instrument triage and showed its feasibility to prioritize instruments for the physical function domain. This process as illustrated in Figure 1 was drafted following a consensus effort from the steering committee including input from PRPs and may be used as a template in guiding subsequent working groups to choose

instruments with high potential for fulfilling the OMERACT Filter 2.1 for instrument selection. Its application may be especially useful when assessing domains that have numerous existing measurement instruments developed over the years, often for other indications, particularly for domains such as physical function and HRQoL. This template may be less useful for highly specific PsA domains such as enthesitis where few instruments are specifically developed and available, so that the working group may not need a method to shortlist instruments.

The work processes in this template (Figure 1) consisted of forming a working group with representative stakeholders, identification of PROMs through SLR, thorough discussions on content and feasibility of the instruments, and achievement of consensus through Delphi exercises. This template provided a platform for the working group to exclude instruments that have inadequate domain match, poor feasibility or otherwise low potential from further formal appraisal using the OMERACT Filter 2.1. It also allowed new instruments that have less established evidence but high potential to be considered for further evaluation.

The strengths of this current report include collaborative work from health care professionals and PRPs to prioritize instruments for further appraisal. The working group members have expertise in the physical function domain in PsA with international representation. There are some limitations in interpretation that require highlighting. The consensus Delphi exercises were conducted among the 13 working group members rather than involving a larger number of stakeholders, recognizing that the discussions among the stakeholders were deep and thorough. During the Delphi exercises, working group members voted for the PROMs based on their overall impression of the PROMs. These gaps will be bridged eventually as each of the prioritized PROMs will be taken forward to formal appraisal using the OMERACT Filter 2.1. Evidence supporting each PROM in the final

standized outcome measurement set will be presented instrument by instrument, and endorsement from a larger GRAPPA and OMERACT community will be sought.

In summary, we report application of consensus-driven template to prioritize multiple instruments for further appraisal for the physical function domain in PsA, in a project to standardize the Core Outcome Set in PsA. We prioritized 6 PROMs for use in RCTs and LOS via a concerted effort from experts and PRPs. These prioritized physical function PROMs will undergo further appraisal using the OMERACT Filter 2.1.

Figure 1. The simple 3 steps to shortlist instruments for a domain

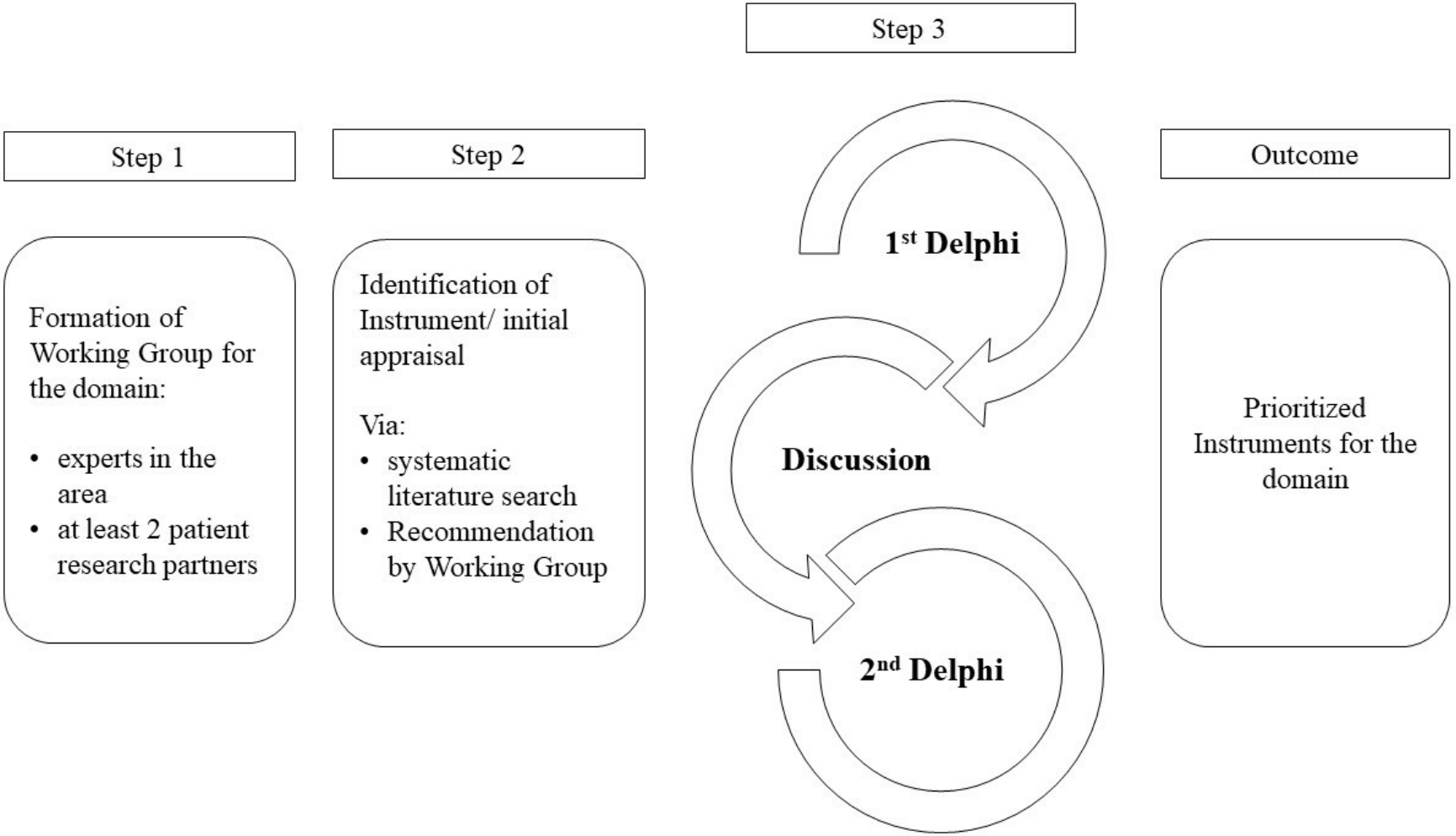

Table 1. Comments from the working group given for each physical function PROM.

PROMs	First Delphi exercise voting results N (%) for "Yes"	For	Against
HAQ-DI	13 (100)	• It has been used in most LOS and RCTs in PsA • Most of the measurement properties have been appraised	• Nil
HAQ-S	9 (69.2)	• The additional item addressed physical impairment related to cervical spine involvement in PsA. • One paper suggested that HAQ-S and HAQ-DI provided similar information. It is possible that there was an inadequate number of patients of each subtype to show the differences, or patients included were not reflective of the full spectrum of axial involvement. • It has been incorporated in the Corrona registry with a larger proportion of PsA patients with axial involvement. Further data analysis may provide an answer to whether it adds new information.	• The additional items (eg, working at a desk, driving a car) are too specific and not relevant for all patients. • It provides no additional information compared to the HAQ-DI.
mHAQ	10 (76.9)	• It is a shorter version of HAQ-DI. • It has been incorporated in the Corrona registry with a larger proportion of PsA patients with axial involvement. Further data analysis may provide an answer to whether it adds new information.	• It may be too brief. • It provides the same information as the HAQ-DI. • There are currently minimal data on its measurement properties
RAPID3	9 (69.2)	• The first 10 items of RAPID3 are actually the MDHAQ, which can be calculated as a Physical	• RAPID3 measures HRQoL. It does not entirely match with the physical function domain.

		Function score.	• Items 1-10 describe physical function, while the rest were pain, patient global assessment and psychological impact. It is not clear if it measure disease activity or impact. • The score categories are confusing (eg, near remission, low severity).
SF-PF10	13 (100)	• The SF-36 has been used in most RCTs for PsA, for which SF-PF10 can be derived.	• Nil
SF-36 PCS	8 (61.5)	• The SF-36 has been used in most RCTs for PsA, and the SF-36 PCS results have been reported in many RCTs.	• The SF-36 PCS is not a measure of physical function; it is calculated based on all 8 domains using a very complicated equation. It measures many concepts in addition to physical function. It is used to determine statistical significance so that the individual domains may be interrogated without a p value correction. • The SF-36 PCS does not match the domain of physical function. It is a measure of HRQoL (includes all 8 weighted domains of the SF-36 questionnaire).
SF-12 PCS	5 (38.5)	• The SF-12 is a shorter version of the SF-36, which may be more feasible than the SF-36.	• Similar to the SF-36, the SF-12 PCS does not match to the domain of physical function, but a measurement of HRQoL. • There are no data for use of SF-12 PCS in PsA. • The SF-12 PCS was not listed in the previous SLR and not listed in the evidence summary table.
PROMIS-PF10a	12 (92.3)	• The PROMIS-PF10a was derived from a huge item bank, and may have higher precision in measurement of physical function.	• The measurement properties of PROMIS-PF10a have not been evaluated in PsA. • It has not been used in any RCT or LOS of PsA.

PsAID functional capacity	11 (84.6)	• The PsAID has received provisional endorsement from OMERACT as a measure of HRQoL in PsA.	• The PsAID should be taken as a whole for the measurement of HRQoL in PsA, rather than broken down into components. • It is an 11-point numeric rating scale for physical function. There is lack of granularity as a single item to measure a domain. The precision is expected to be low.
AIMS	4 (30.8)	• It seems to be thorough and have good domain match with the qualitative description (arm function, mobility, walking and bending, hand and finger).	• It is too long to be feasible. • It has not been used for many years. • Patients' previous feedback with AIMS was negative. It would be difficult to persuade patients to complete PROMs they do not like. • There are only limited data available on measurement properties.
BASFI	8 (61.5)	• It has relevant items for axial function including the cervical spine.	• It is not meant to measure physical function in PsA. There is a lack of content validity in measuring physical function in PsA. • The content does not represent concerns in PsA patients with axial involvement. It is not specific to PsA patients with axial involvement. • It has too much focus on axial function. • It has poor psychometric properties in PsA. • It gives the same information as the HAQ-DI.
ACR functional class	4 (30.8)	• While developed for RA, it has some broadly generalizable information usable in clinical trials, such as that for inclusion or exclusion criteria.	• It is too brief. • It is an outdated instrument that is not in use. • It may lack content for PsA patients nowadays where severe disabling is seldomly seen. • The level of response and categories are difficult to understand. • It is not a PROM to measure the perceived

			physical function from patients' perspective. • It is too crude, only having a few levels of responses that span across fully functional to bedridden. • Given the crude categories, the responsiveness is expected to be poor.

Abbreviations: ACR, American College of Rheumatology; AIMS, Arthritis Impact Measurement Scales; BASFI, Bath Ankylosing Spondylitis Functional index; HAQ, Health Assessment Questionnaire (HAQ-S: Spondyloarthropathy, HAQ-DI: Disability Index); mHAQ; modified HAQ; MDHAQ, multidimensional HAQ; SF-36, Medical Outcomes Study 36-item Short Form Survey (PCS: Physical Component Summary; PF: SF-36 physical functioning domain); PsAID, Psoriatic Arthritis Impact of Disease; PROMs, patient-reported outcome measures; PROMIS, Patient-Reported Outcomes Measurement Information System.

Table 2. Results of the two rounds of Delphi exercise

PROMs for physical function	First Delphi exercise voting results N (%) for "Yes"	Consensus questions developed for second Delphi	Second Delphi Exercise Voting results N (%) for "Yes" [final decision]
HAQ-DI	13 (100)	• HAQ-DI received 100% votes in the first Delphi. • Do you think we should take HAQ-DI to appraisal via OMERACT Filter 2.1? Yes/ No	13 (100) [included]
HAQ-S	9 (69.2)	• HAQ-S has 5 additional items for spine added to HAQ-DI. It was previously shown to give similar information as HAQ-DI. However, it may be relevant for patients with axial PsA. It has been incorporated in the Corrona registry with data pending. • HAQ-S received 69.2% votes in first Delphi. • Given this consideration, should we appraise HAQ-S via OMERACT Filter 2.1? Yes/ No	10 (76.9) [included]
HAQ-DI and HAQ-S		• Secondly, are you agreeable to see HAQ-DI and HAQ-S as a family. If evidence is supportive of HAQ-S as useful for axial PsA, to allow using either of the HAQ for trials for different purposes? Yes/No	11 (84.6) [included]
mHAQ	10 (76.9)	• mHAQ is a shorter version of HAQ-DI (8-items) • It received 76.9% votes in the first Delphi. • Do you think we should appraise mHAQ via OMERACT Filter 2.1? Yes/ No	11 (84.6) [included]

MDHAQ	Voted under RAPID3 9 (69.2)	• MDHAQ is modified from HAQ. It consists of a 10-item physical function score, pain, stiffness, fatigue, and patient global. • Rated under RAPID3 (which consisted of the 10-item physical function, pain and patient global), it received a 69.2% vote in the first Delphi. • The 10-item physical function of MDHAQ is purely for physical function and can be taken as independent scale. • Do you think we should appraise the physical function score of MDHAQ via OMERACT Filter 2.1? Yes/ No	10 (76.9) [included]
SF-36 PF10	13 (100)	• SF-PF10 has received a 100% vote in the first Delphi. • Do you think we should take SF-PF10 to appraisal via OMERACT Filter 2.1? Yes/ No	13 (100) [included]
SF-36 PCS	8 (61.5)	• SF-36 PCS has been reported in clinical trials. • However, it is not measuring the domain of physical function. • It received a 61.5% vote in the first Delphi. • Given this consideration, should we appraise SF36 PCS via OMERACT filter 2.1? Yes/ No	2 (15.4) [excluded]
SF-12 PCS	5 (38.5)	• SF-12 PCS was not in the systematic review. There is no study that evaluates its use in PsA. It is excluded for further voting.	NA [excluded]
PROMIS-PF10a	12 (92.3)	• PROMIS-PF10a (short form) has only 10 items. • It is a promising generic measure of physical function • It received a 92.3% vote in the first Delphi • Do you think we should appraise the PROMIS-PF10a via OMERACT Filter 2.1? Yes/ No	13 (100) [included]

PsAID item 5 functional capacity	11 (84.6)	• PsAID item #5 functional capacity received 84.6% votes in the first Delphi. • Discussion has been not to select individual items from an instrument, single items do not measure a domain well, there has been no validation of the PsAID item #5 as a stand-alone measure of physical function, and PsAID12 as a whole does not match the physical function domain. • It may be relevant to see if #5 functional capacity is consistent with other physical function measures. • Given this consideration, should PsAID #5 functional capacity be appraised via OMERACT Filter 2.1? Yes/ No	4 (30.8) [excluded]
AIMS	4 (30.8)	• AIMS is a long instrument and lacks feasibility. It has not been used in the community. It has received only 30.8% vote in the first Delphi. • Discussion around BASFI has been on lack of domain match, even for axial PsA; and giving similar information as HAQ-DI or HAQ-S. It received a 61.5% vote in the first Delphi. • Discussion on ACR functional class has been that it is too crude, lacks domain match with lesser physical impairments among patients nowadays, and is not used much in the field. It has received only 30.8% vote in the first Delphi. • Given the above considerations, should AIMS, BASFI and ACR functional class be appraised via OMERACT Filter 2.1? Yes/ No	0 (0) [excluded]
BASFI	8 (61.5)		
ACR functional class	4 (30.8)		

Response rate from 13 working group members 100%.

Abbreviations: ACR, American College of Rheumatology; AIMS, Arthritis Impact Measurement Scales; BASFI, Bath Ankylosing Spondylitis Functional index; HAQ, Health Assessment Questionnaire (HAQ-S: Spondyloarthropathy, HAQ-DI: Disability Index); mHAQ; modified HAQ; MDHAQ, multidimensional HAQ; SF-36, Medical Outcome Survey Short Form 36-item Health Survey (PCS: Physical Component Summary; PF: SF-36 physical function subscale); PsAID, Psoriatic Arthritis Impact of Disease; PROMs, patient-reported outcome measures; PROMIS, Patient-Reported Outcomes Measurement Information System.

Supplementary Table 1. Examples of quotations of Physical Function taken from the International focus group study by GRAPPA

- *What affects me the most is not being able to do certain things because we get immobilized. I can't bend down, I can't lift my arm much, because it hurts, so there are certain things you can't do. (Brazil)*
- *[The walking cane] helps me to walk and most importantly it helps me keep my balance. It didn't really change my relationship with the people around me. It is to help me, because my ankle, my knee, my hip, suddenly... I lose control. I recover by using my cane, still suffering, as my hands are also painful. (France)*
- *It is hard to move or lift objects and once in a while it is even hard to wake up and get perpendicular. There are days that waking up and getting up is really difficult, even turning around in bed can be difficult. (France)*
- *-Sometimes I cannot even peel potatoes, well then I'm working with an electric potato peeler (laughs)-That goes electrically because then you cannot hold it.-Yes the housekeeping is pretty restrictive. (exchange between participants, Netherlands)*
- *I'm just so stiff. It's hard to straighten by body out and get my legs under me and actually be able to walk out of the room. (USA)*

Reference

1. Gladman DD, Antoni C, Mease P, Clegg DO, Nash P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. Ann Rheum Dis. 2005;64 Suppl 2:ii14-7.
2. Ritchlin CT, Colbert RA, Gladman DD. Psoriatic Arthritis. N Engl J Med. 2017;376(21):2095-6.
3. Orbai AM, de Wit M, Mease P, Shea JA, Gossec L, Leung YY, et al. International patient and physician consensus on a psoriatic arthritis core outcome set for clinical trials. Ann Rheum Dis. 2017;76(4):673-80.
4. Bissonnette R, Luger T, Thaci D, Toth D, Lacombe A, Xia S, et al. Secukinumab demonstrates high sustained efficacy and a favourable safety profile in patients with moderate-to-severe psoriasis through 5 years of treatment (SCULPTURE Extension Study). J Eur Acad Dermatol Venereol. 2018;32(9):1507-14.
5. Boers M, Kirwan JR, Tugwell P, Beaton D, Bingham III CO, Conaghan PG, et al. The OMERACT Handbook. 2017. [Online]. Available at: https://img1.wsimg.com/blobby/go/e6f90123-ada6-4eef-903a-df8fb00eb48d/downloads/1cn7eh07f_621480.pdf. Accessed 26 Sep, 2019.
6. Ramiro S, Smolen JS, Landewe R, Heijde DV, Gossec L. How are enthesitis, dactylitis and nail involvement measured and reported in recent clinical trials of psoriatic arthritis? A systematic literature review. Ann Rheum Dis. 2018;77(5):782-3.
7. Tillett W, Orbai AM, Ogdie A, Leung YY, Strand V, Gladman DD, et al. GRAPPA-OMERACT initiative to standardise outcomes in psoriatic arthritis clinical trials and longitudinal observational studies. Ann Rheum Dis. 2018;77(5):e23.
8. Beaton DE, Maxwell LJ, Shea BJ, Wells GA, Boers M, Grosskleg S, et al. Instrument Selection Using the OMERACT Filter 2.1: The OMERACT Methodology. J Rheumatol. 2019;46(8):1028-35.
9. Orbai AM, de Wit M, Mease PJ, Callis Duffin K, Elmamoun M, Tillett W, et al. Updating the Psoriatic Arthritis (PsA) Core Domain Set: A Report from the PsA Workshop at OMERACT 2016. J Rheumatol. 2017;44(10):1522-8.
10. Stamm TA, Nell V, Mathis M, Coenen M, Aletaha D, Cieza A, et al. Concepts important to patients with psoriatic arthritis are not adequately covered by standard measures of functioning. Arthritis Rheum. 2007;57(3):487-94.
11. Dures E, Hewlett S, Lord J, Bowen C, McHugh N, Group PS, et al. Important Treatment Outcomes for Patients with Psoriatic Arthritis: A Multisite Qualitative Study. Patient. 2017;10(4):455-62.
12. Mease P, Strand V, Gladman D. Functional impairment measurement in psoriatic arthritis: Importance and challenges. Semin Arthritis Rheum. 2018;48(3):436-48.
13. Leung YY, Orbai AM, Ogdie A, Coates LC, de Wit M, Callis Duffin K, et al. The GRAPPA-OMERACT Psoriatic Arthritis Working Group at the 2018 Annual Meeting: Report and Plan for Completing the Core Outcome Measurement Set. The Journal of rheumatology Supplement. 2019;95:33-7.
14. Hojgaard P, Klokker L, Orbai AM, Holmsted K, Bartels EM, Leung YY, et al. A systematic review of measurement properties of patient reported outcome measures in psoriatic arthritis: A GRAPPA-OMERACT initiative. Semin Arthritis Rheum. 2018;47(5):654-65.
15. Prinsen CA, Vohra S, Rose MR, Boers M, Tugwell P, Clarke M, et al. How to select outcome measurement instruments for outcomes included in a "Core Outcome Set" - a practical guideline. Trials. 2016;17(1):449.
16. Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, et al. The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. J Clin Epidemiol. 2010;63(7):737-45.
17. Holland R, Tillett W, Ogdie A, Leung YY, Gladman DD, Callis Duffin K, et al. Content and Face Validity and Feasibility of 5 Candidate Instruments for Psoriatic Arthritis Randomized Controlled

Trials: The PsA OMERACT Core Set Workshop at the GRAPPA 2017 Annual Meeting. The Journal of rheumatology Supplement. 2018;94:17-25.
18. Fries JF, Spitz P, Kraines RG, Holman HR. Measurement of patient outcome in arthritis. Arthritis Rheum. 1980;23(2):137-45.
19. Daltroy LH, Larson MG, Roberts NW, Liang MH. A modification of the Health Assessment Questionnaire for the spondyloarthropathies. J Rheumatol. 1990;17(7):946-50.
20. Pincus T, Summey JA, Soraci SA, Jr., Wallston KA, Hummon NP. Assessment of patient satisfaction in activities of daily living using a modified Stanford Health Assessment Questionnaire. Arthritis Rheum. 1983;26(11):1346-53.
21. Ware JE, Jr., Sherbourne CD. The MOS 36-item short-form health survey (SF-36). I. Conceptual framework and item selection. Med Care. 1992;30(6):473-83.
22. Ware J, Jr., Kosinski M, Keller SD. A 12-Item Short-Form Health Survey: construction of scales and preliminary tests of reliability and validity. Med Care. 1996;34(3):220-33.
23. Gossec L, de Wit M, Kiltz U, Braun J, Kalyoncu U, Scrivo R, et al. A patient-derived and patient-reported outcome measure for assessing psoriatic arthritis: elaboration and preliminary validation of the Psoriatic Arthritis Impact of Disease (PsAID) questionnaire, a 13-country EULAR initiative. Ann Rheum Dis. 2014;73(6):1012-9.
24. Meenan RF, Mason JH, Anderson JJ, Guccione AA, Kazis LE. AIMS2. The content and properties of a revised and expanded Arthritis Impact Measurement Scales Health Status Questionnaire. Arthritis Rheum. 1992;35(1):1-10.
25. Calin A, Garrett S, Whitelock H, Kennedy LG, O'Hea J, Mallorie P, et al. A new approach to defining functional ability in ankylosing spondylitis: the development of the Bath Ankylosing Spondylitis Functional Index. J Rheumatol. 1994;21(12):2281-5.
26. Hochberg MC, Chang RW, Dwosh I, Lindsey S, Pincus T, Wolfe F. The American College of Rheumatology 1991 revised criteria for the classification of global functional status in rheumatoid arthritis. Arthritis Rheum. 1992;35(5):498-502.
27. Pincus T, Swearingen C, Wolfe F. Toward a multidimensional Health Assessment Questionnaire (MDHAQ): assessment of advanced activities of daily living and psychological status in the patient-friendly health assessment questionnaire format. Arthritis Rheum. 1999;42(10):2220-30.
28. Rose M, Bjorner JB, Gandek B, Bruce B, Fries JF, Ware JE, Jr. The PROMIS Physical Function item bank was calibrated to a standardized metric and shown to improve measurement efficiency. J Clin Epidemiol. 2014;67(5):516-26.
29. Pincus T, Yazici Y, Bergman MJ. RAPID3, an index to assess and monitor patients with rheumatoid arthritis, without formal joint counts: similar results to DAS28 and CDAI in clinical trials and clinical care. Rheum Dis Clin North Am. 2009;35(4):773-8, viii.
30. Blackmore MG, Gladman DD, Husted J, Long JA, Farewell VT. Measuring health status in psoriatic arthritis: the Health Assessment Questionnaire and its modification. J Rheumatol. 1995;22(5):886-93.
31. Ware JE, Kosinski M, Keller SD. SF-36 Physical and Mental Health Summary Scales: a User's Manual. Boston, MA: Health Assessment Lab, New England Medical Center; 1994.
32. Allard A, Antony A, Shaddick G, Jadon DR, Cavill C, Robinson G, et al. Trajectory of radiographic change over a decade: the effect of transition from conventional synthetic disease-modifying antirheumatic drugs to anti-tumour necrosis factor in patients with psoriatic arthritis. Rheumatology (Oxford, England). 2019;58(2):269-73.